DOCUMENT RESUME

ED 326 478 SO 030 447

AUTHOR Gehshan, Shelly; And Others
TITLE College Admissions Tests: Opportunities or Roadblocks?
INSTITUTION American Association of Univ. Women, Washington, D.C.
PUB DATE Jun 88
NOTE 6p.; For related documents, see SO 030 445-446.
AVAILABLE FROM American Association of University Women, Sales Office, 11722 Parklawn Drive, Rockville, MD 20852 (1.00).
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Admission Criteria; College Admission; *College Entrance Examinations; Educational Policy; *Equal Education; *Females; Higher Education; Males; Sex Bias; Sex Differences; *Sex Discrimination; *Standardized Tests; Womens Studies
IDENTIFIERS *Scholastic Aptitude Test

ABSTRACT

Recent research shows that the Scholastic Aptitude Test (SAT) systematically underestimates girls' ability to succeed in college, and suggests that this is because of a subtle pervasive bias in the construction of this and most other standardized tests. This issue brief reviews the research on gender bias in college admissions testing and the impact of test bias on girls, discusses the function of testing in U.S. society, and suggests some alternatives for addressing this problem. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED326478

College Admissions Tests: Opportunities or Roadblocks

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

KATHY VANDELL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SO 030 447

AMERICAN ASSOCIATION

OF UNIVERSITY WOMEN

June 1988

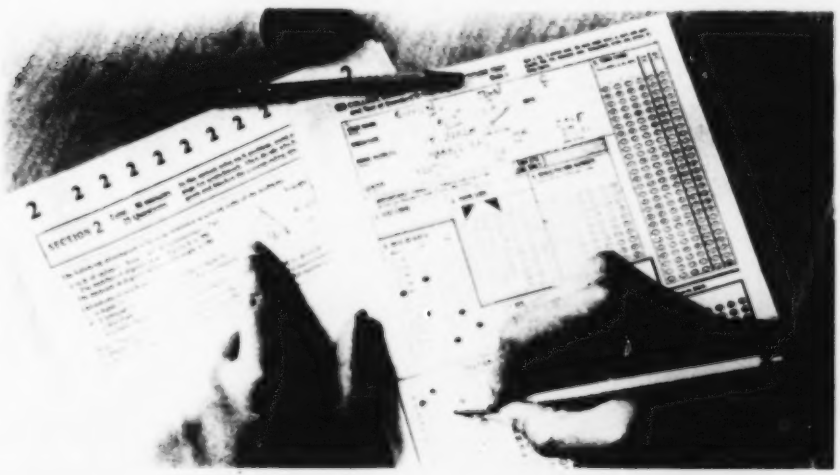
College Admissions Tests: Opportunities or Roadblocks?

WHY KAREN HAS FEWER CHOICES

Mark and Karen are high school seniors. Both rank in the top 10 percent of their class. Both are applying to several of the more than 1,500—88 percent of—U.S. colleges that require applicants to submit Scholastic Aptitude Test (SAT) scores.

Once the test is over, however, the similarities end, because if their test scores reflect the national average, Karen's overall score will be 60 points lower than Mark's. As a result, Karen's predicted grade-point average for her freshman year of college might well fall below the cut-off for admission to the most selective colleges, while Mark's might fall above it. Karen would get into college—but she would be offered fewer, and perhaps less desirable, choices. She would also be less likely to win one of the hundreds of scholarships awarded largely on the basis of SAT scores. Perhaps most important, Karen might lower her academic and professional goals because of her scores. And all this will happen despite the fact that Karen's actual freshman grades will not only be higher than her test scores predict, they'll be higher than Mark's.

Recent research shows that the SAT systematically underestimates girls' ability to succeed in college, and suggests that this is because of a subtle, pervasive bias in the construction of this and most other standardized tests. This brief will review the research on gender bias in college admissions testing and the impact of test bias on girls, discuss the function of testing in our society, and suggest some alternatives for addressing this problem.



WHERE THE TESTS FAIL

In 1987, boys scored an average of 57 points higher than girls on the most common college admissions test, the SAT. SAT scores are reported for two areas of ability, verbal and mathematical. In 1987, boys' average scores were 47 points higher than girls' on the mathematical section, and 10 points higher on the verbal section.[1]

Boys' average scores are also higher than girls' on the American College Assessment Test (ACT), although the gap is smaller than on the SAT. ACT scores are reported for four subject areas: mathematics, natural sciences, social studies, and English. In 1987, boys outscored girls on every portion of the ACT except the English section.[2]

Although girls' scores are lower than boys' on both tests, current research is concentrated on the SAT, primarily because the SAT affects more students. It is taken by 1.6 million students annually, whereas one million take the ACT.[3] In addition, the SAT is currently the only standardized test on which girls receive lower verbal scores than boys. Although boys have always outscored girls on the mathematical section, girls scored higher than boys on the verbal section of the SAT until 1972.[4] The reversal of this pattern has drawn attention to the issue of gender bias in testing, and focused concern on the SAT.

Although the Educational Testing Service (ETS), the non-profit corporation that produces the SAT, has been studying the difference between boys' and girls' scores for over 20 years, the first ETS study that suggested the test was biased against girls was published in 1984. The SAT is designed to predict the grades students

If the SAT were an accurate predictor of students' freshman grades, girls would score 20 points higher than boys, not 57 points lower.

will receive in their freshman year of college. If the test is valid, students with higher SAT scores should earn higher grades during their first year of college. In the 1984 study, however, ETS found that girls' actual freshman grades were higher, and boys' lower, than predicted by their SAT scores.[5] The test therefore is not an accurate predictor of college performance, and admissions decisions based on SAT scores discriminate against girls.

If the SAT predicted equally well for both sexes, girls would score about 20 points higher than boys.[6] The error in grade prediction is even higher for reentry women, according to an earlier ETS study showing that older women students had lower SATs and higher grades than both their male counterparts and younger, traditional students.[7]

WHY ARE GIRLS' SCORES LOWER?

Recent research suggests that the gap between girls' and boys' test scores stems from the way the test is constructed. Comparisons of girls' and boys' performance on individual test questions reveal that girls excel on certain types of questions, and boys on others.

While the reasons for these differences are not always clear, they are usually related to gender-stereotyped socialization—including differences in teachers' and parents' expectations and selective exposure to various toys, role models, and activities—that encourage girls and boys to develop different interests, skills, and approaches to problems. For example, ETS has known since 1965 that boys tend to perform better on the questions it classifies as "practical affairs" and "science-related," while girls do better on "aesthetic/philosophical" and "human relations" questions.[8]

The bias arises because the types of questions favoring each gender are not evenly balanced on the SAT. In fact, ETS changed the composition of the verbal test in the early 1960s to put greater emphasis on science, and the decline in girls' verbal SAT scores dates from that decision.[9] Girls are doing increasingly poorly on the SAT because they are, in effect, being tested for maleness.

This is reinforced by the portrayal, or lack thereof, of women in test questions. An analysis of 24 passages from the reading comprehension sections of the verbal SAT given in 1984–1985 contained 34 references to famous men, and mentioned eight other male characters by name. The only famous woman mentioned was Margaret Mead, in a passage criticizing her work, and two other unnamed female characters were mentioned. This presents a subtle but strong message of male achievement and female failure and anonymity to girls (and boys) taking the test. It affects girls' scores, since research shows that girls are more likely to get the right answer to questions containing references to more women than men, or an equal number of women and men, or in which sex is unspecified.[10]

Culturally biased test questions are the source of another gap in scores. Minority and low-income girls suffer doubly at the hands of the SAT as a result. In 1985, black girls scored an average of 207 points lower than white girls, and students with family incomes of between $12,000 and $18,000 scored an average of 145 points lower than students with family incomes of above $50,000.[11] A 1987 examination of the SAT by the National Center for Fair & Open Testing (FairTest) found many questions that required students to be familiar with the activities and vocabulary of upper-middle-class Americans such as golf, tennis, pirouettes, property taxes, kettledrums, minuets, melodeons, tympanists, polo, and horseback riding.[12]

Overwhelming evidence of gender and cultural bias in the SAT also belies ETS's claims that it is an aptitude test, which measures ability, rather than an achievement test, which measures knowledge. Were it a valid aptitude test, factors such as family income, previous educational environment, and gender would have little effect on scores.

An analysis of 24 passages from the SAT's reading comprehension sections contained 34 references to famous men, but only one reference to a famous woman—Margaret Mead—and that was in a passage criticizing her work.

TESTING THE TESTERS

ETS and its parent company, the College Board, claim that the gap in boys' and girls' SAT scores results primarily from differences in high school preparation and socioeconomic factors. Girls now comprise 52 percent of those taking the SAT—up from 44 percent in 1964—and many of the new test-takers are girls from low-income and minority families, who typically have less access to quality educational opportunities. ETS and the College Board say these new test-takers are responsible for the gap in scores.[13] But these claims are countered by evidence that the gender gap persists within minority groups where girls are not the majority of test-takers. In 1985, an equal number of Mexican-American boys and girls took the SAT, and boys scored an average of 70 points higher; in the same year, more Asian-American boys than girls took the test, and the gap was 49 points.[14]

Although ETS has occasionally conducted studies analyzing the performance of girls and minorities on the SAT, they made no effort to balance the test until recently, and current procedures are inadequate to eliminate this bias. ETS "tests the test" in two ways: subjective reviews by ETS staff and advisory committees, and statistical analyses of test results. Each question is pretested in a special, non-scored section of at least one SAT. If more students with high overall scores get a question wrong than students with lower scores, the question is returned to the review committees for further examination; the committee may or may not discard the question.[15] This procedure assumes that the overall test score is an accurate measure of ability; i.e., that the standard for accuracy of a question is its correlation with performance on the test as a whole. If the test is biased, so is the review process.

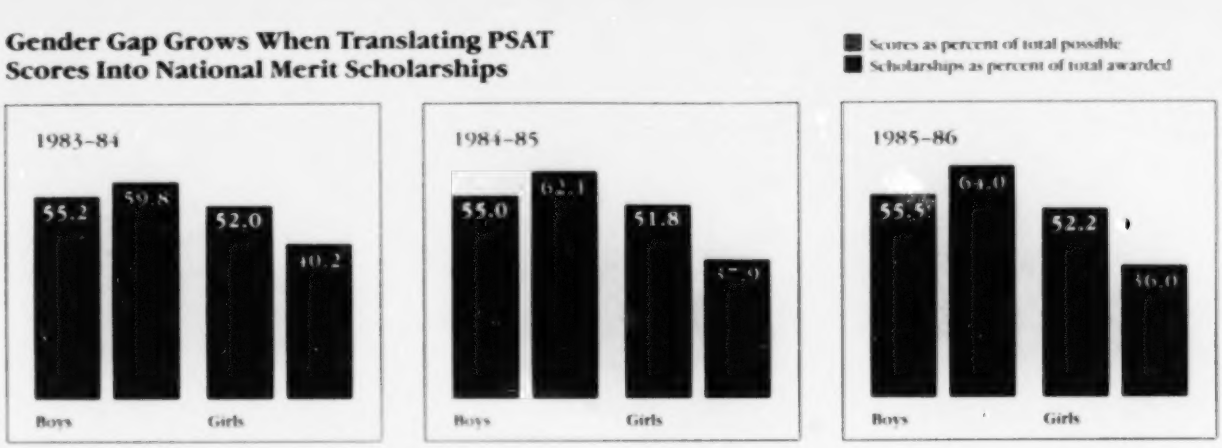
ETS has responded to recent charges of test bias by adding to its subjective reviews a "sensitivity review," which attempts to identify inappropriate language in test questions. Since 1982, ETS has also conducted research on differential item functioning (DIF), a statistical comparison of the performance of girls and boys and various racial and ethnic groups on individual test items. If the percentage of students answering correctly is significantly higher for one group, the question is flagged for further review. Within the last year ETS has begun to move from research to implementation of routine DIF analyses of every pretested question.[16]

FairTest questions whether this represents real reform or a public relations strategy. ETS's DIF analysis procedure has two flaws. ETS does not necessarily discard test items that are flagged by DIF; they only return them for further review. In the past, ETS has been reluctant to discard questions due to the high cost of pretesting. Further, the DIF procedure is inadequate. ETS only compares the individual item performance of girls and boys who have equal overall scores on the test. Although they justify this as a way of holding constant for preparation and other factors that affect scores, it assumes overall scores to be an unbiased standard of performance. In other words, ETS is testing for bias in each question by assuming the test itself to be unbiased.[17]

"SAT" MEANS "SCHOLARSHIPS AREN'T THERE"

Eighty-eight percent of America's colleges and universities use the SAT to predict students' success in college, usually in combination with either high school grades or high school class rank. Some colleges feed SAT scores into computer programs that estimate freshman year grade-point averages, as ETS recommends. Others use a scale of test scores and high school grade-point averages or class ranks to establish admissions cutoffs. Interviews, recommendations, participation in school and community activities, and other criteria, are also con-

Gender Gap Grows When Translating PSAT Scores Into National Merit Scholarships



SOURCE: FairTest

sidered by most colleges, in keeping with the College Board's guidelines for use of the SAT. Test scores are the sole or primary admissions criterion for 20 percent of the nation's 1,900 four-year colleges.[18] Because SAT scores underestimate their academic performance, girls are more likely to be denied admission to colleges that use the SAT in admissions decisions. The extent of this discrimination depends on the relative importance of the test scores in admissions decisions.

In addition, students often decide not to apply to particular colleges, or not to apply to any college, when their test scores are lower than the published requirements or averages of the colleges they are considering. If women's test scores are systematically lower than they should be, not only will they be unfairly denied admission to some colleges, but most will choose not to apply to colleges with higher SAT score requirements. Some colleges are trying to compensate for such problems by adjusting for girls' scores, formally or informally, but girls taking the test don't know that. Adjusting for potential bias in scores doesn't help students if they are discouraged from applying in the first place.[19]

Lower test scores also translate into fewer and less lucrative scholarships for women. Semifinalists in the contest for National Merit Scholarships, for example, are chosen solely on the basis of their Preliminary-SAT (PSAT) scores taken in their junior year of high school. Girls averaged 53 points lower than boys on the PSAT in 1985–86, and received only 36 percent of these scholarships, down from 40 percent in 1983–84 (see above).[20]

Many states and private organizations also offer scholarship programs based largely on test scores. In 1986–87, 72 percent of New York's Empire State Scholarships, then based solely on SAT scores, went to boys. These scholarships provide $10,000 over four years. When a state law requiring that high school grades be given equal weight with test scores was passed in 1987, the number of Empire State Scholarships received by girls jumped from 28 to 40 percent, a 43 percent increase.[21]

Women's aspirations for achievement may also be adversely affected by biased tests. Abundant research suggests that stereotypes about the abilities and appropriate roles for men and women could decrease girls' confidence in their intellectual abilities, lower their expectations of success, and alter their educational and career choices. Because women typically attribute their successes to effort and their failures to lack of ability, they tend to interpret their test scores as a judgment of their abilities.[22] In addition to the psychological damage this does to each individual, it results in a serious loss of human potential.

In 1986–87, when SAT scores were the sole criteria, only 28 percent of New York's Empire State Scholarships went to girls. In 1987–88, when high school grades were given equal weight with SAT scores, girls received 40 percent of the scholarships.

GUARDING THE GATES TO OPPORTUNITY

The role of standardized tests in American society has increased consistently since the Army's adoption of intelligence testing in 1917, but has always been controversial. Increased reliance on testing has historically been associated with periods of high immigration and associated fears of disintegration of the national culture. Columbia University's first use of an admissions test in 1918, for instance, was an attempt to stem the influx of Jewish students. Because many educators of the time believed that northern European ethnic groups were genetically superior in intelligence, they saw no contradiction in asserting that the tests would measure merit scientifically and exclude people from certain backgrounds from intellectual and professional institutions.[23]

Tests now serve as the gatekeepers to greater opportunity at every level of society. Americans take standardized tests to determine their intelligence and progress in the three R's in elementary school, to determine vocational interest and aptitude, to gain admission to college and professional schools, and to qualify for licenses in professions ranging from real estate to medicine. The SAT was first administered in 1926, but was not widely used in college admissions testing until 1947, when ETS was established to administer it. Although ETS is a nonprofit corporation, it is not quite the disinterested scholarly institution many perceive it to be. In 1976, *Forbes* described it as "one of the hottest little growth companies in U.S. business."[24]

Criticism of standardized testing in general, and ETS in particular, has been growing since the mid-1960s, when accusations of racial bias first called attention to tests' validity. Consumer activists also began to question the unparalleled power of ETS, a company which affects thousands of lives and is largely unaccountable for its actions. Every SAT question is pretested on an actual test, for instance, as part of an experimental section which is not identified, and which students cannot choose not to take. At most universities, such practices would violate human experimentation ethics rules. Few students would agree with ETS's contention that they can choose whether to take the test, when so many college admissions offices require it.

Test consumers have begun to demand more accountability. In 1979, New York State passed the first truth-in-testing law despite heavy lobbying by ETS. The law required ETS to return corrected answer sheets and examination copies to students upon request. With the veil of secrecy lifted, ETS extended this return-upon-

When the University of North Carolina at Chapel Hill discovered that its fall 1986 freshman class was 60 percent women, some trustees suggested raising SAT score cut-offs to exclude more women. If the trend continued, one trustee asserted, "I think the amount of alumni giving would go down. I also think the influence of alumni in the legislature would go down."

request practice to all students in 1981. Nearly everything that is now known about bias on the test is the result of this new access.

In 1984, ETS reached an out-of-court settlement of a lawsuit challenging its insurance licensing test. The settlement specified a procedure for determining racial bias: if the average scores for blacks are more than 15 percent different from the scores for whites on a particular question, that question will be discarded on future exams. This procedure, known as the Golden Rule procedure, could be applied to gender bias as well, but ETS has not done so, and has preferred to use the questionable DIF procedures described above and to reserve the right to retain biased questions.

A package of testing oversight laws passed last year in New York provides a model for activists. This legislation required ETS to publish a statistical report of the performance of women, men, and minorities on the test, established a commission to review all postsecondary school examination testing for bias, and required the state to weigh equally high school grades and the SAT in awarding scholarships.

WHAT CAN BE DONE?

Activists can

- Share the information in this brief with students, parents, high schools, and colleges.
- Educate young women, parents, and colleges about the limitations of the test. Young women should understand that their ability is not being measured by the SAT, and be encouraged to discuss other indicators of college success with admissions officials.
- Ask colleges to evaluate their admissions criteria. Is the SAT predicting women students' success accurately? How are admissions affected by use of the SAT? How is the distribution of financial aid being affected by test scores?
- Ask colleges to collect and publish statistics on admissions and scholarships by gender, race, ethnicity, and income.
- Demand a fair test from the testing companies.

Colleges can

- Evaluate and publish studies of how test scores affect their admission of girls and minorities.
- Diversify their admissions and scholarship criteria. Colleges should rely more on writing samples, personal interviews, and evaluation of community activities in addition to students' academic records.

State governments can

- Examine the distribution of state scholarships and the affect of test scores on their distribution.
- Follow New York State's example and set up panels to study admissions tests and recommend appropriate legislation.
- Require state universities to evaluate and publish studies of their admissions programs, and diversify their admissions criteria.
- Require tests given in the state to use the Golden Rule procedure described above.

ETS can

- Fix the test. Replace the DIF procedure with the Golden Rule procedure or develop an alternative test for gender bias that is not tied to overall scores, and make a commitment to discard all biased test questions.

FOOTNOTES

1. Admission Testing Program of the College Board, *National Report: College Bound Seniors, 1987*, p. v.
2. American College Assessment Test staff, telephone conversation with AAUW Program and Policy Department Research Associate Kathy Vandell, April 8, 1988.
3. Ernest L. Boyer, *College: The Undergraduate Experience in America*, The Carnegie Foundation for the Advancement of Teaching, (New York: Harper & Row, 1987), p. 33.
4. Phyllis Rosser, "Girls, Boys, and the SAT: Can We Even the Score," *NEA Today*, vol. 6 (January 1988), p. 48, 50.
5. Mary Jo Clark and Jerilee Grandy, *Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers*, College Board Report 84-8 (New York: CEEB, 1984), pp. 20–23.
6. Phyllis Rosser, *Sex Bias in College Admissions Tests: Why Women Lose Out*, (Boston: FairTest, 1987), p. 2.
7. Patricia Lund Casserly, *Older Students and the SAT*, College Board Report 82-8 (New York: College Entrance Examination Board, 1982), pp. 3–5.
8. Thomas F. Donlon and William H. Angoff, "Scholastic Aptitude Test," Chapter 2 in William Angoff, ed., *The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests* (New York: CEEB, 1971), p. 26.
9. Rosser, March 8, 1988, and Maureen Welsh, The College Board, April 7, 1988, telephone conversations with Vandell.
10. Rosser, *Sex Bias*, op. cit., Appendix 5.
11. Leonard Ramist and Solomon Arbeiter, *Profiles, College-Bound Seniors, 1985* (New York: CEEB, 1986), pp. 12–97.
12. Rosser, *Sex Bias*, p. 16.
13. Gretchen Rigol, *College Board News*, Summer 1987.
14. *Profiles 1985*, pp. 12–97.
15. Angoff, *op cit.*, Thomas F. Donlon, ed., *The College Board Handbook for the Scholastic Aptitude Test and Achievement Tests* (New York: CEEB, 1984); Welsh, April 7, 1988, and Cathy Wendler, Educational Testing Service, April 15, 1988, telephone conversations with Vandell.
16. Donlon, Chapters 7 and 10; Welsh and Wendler, op cit, telephone conversations with Vandell.
17. FairTest Examiner, Spring 1988, p. 14.
18. Ann Grosso, The College Board, quoted in Causewell Vaughan, "SATs Harm: (A) Women & (B) Minorities," New York *Daily News*, November 29, 1987.
19. Jonathon Rowe, "Scholastic Aptitude Test Scores: Are They Unfair to Women?" *Christian Science Monitor*, April 20, 1987.
20. Rosser, *Sex Bias*, op. cit., pp. 3, 9.
21. *Chronicle of Higher Education*, April 6, 1988, p. A-27.
22. Jacquelynne S. Eccles, "Gender Roles and Women's Achievement-Related Decisions," *Psychology of Women Quarterly* 11 (1987): 135–172; Eleanor E. Maccoby and Carol Nagy Jacklin, *The Psychology of Sex Differences* (Stanford: Stanford University Press, 1974).
23. James Crouse and Dale Trusheim, *The Case Against the SAT* (Chicago: University of Chicago Press, 1988), Chapter 2.
24. *Forbes*, November 15, 1976, p. 89.
25. Raleigh, NC, *News and Observer*, Dec. 27, 1986, p. 1A.

American Association of University Women
2401 Virginia Avenue, N.W.
Washington, DC 20037

AAUW promotes equity for women, education and self-development over the life span, and positive societal change.

The AAUW Educational Foundation provides funds to advance education, research, and self-development for women and to foster equity and positive societal change.

Sarah Harder, President, AAUW; Chair of the Board, AAUW Educational Foundation

Mary A. Grefe, President, AAUW Educational Foundation

Anne L. Bryant, Ed.D., Executive Director

Ann Chipley, Director, Program and Policy Department

This brief was prepared by Shelly Gehshan, Assistant Director, and Kathy Vandell, Research Associate, Program and Policy Department, and Jonathan Kronstadt, Associate Editor, Communications Department. Single copies ($1 for postage and handling for AAUW members, $2 for nonmembers) and bulk orders ($1.50/2–10, $1/11–100, $.80/101 and over) are available from the AAUW Sales Office, 2401 Virginia Avenue, N.W., Washington, DC 20037.

6/88–12,000